KNOWLEDGE DRIVEN IMAGE MINING WITH MIXTURE DENSITY MERCER KERNELS

ASHOK N. SRIVASTAVA, PH.D. AND NIKUNJ OZA, PH.D.

Manuscript for the ESA-EUSC 2004: Workshop on the Theory and Applications of Knowledge Driven Image Information Mining, with focus on Earth Observation

Abstract. This paper presents a new methodology for automatic knowledge driven image mining based on the theory of Mercer Kernels, which are highly nonlinear symmetric positive definite mappings from the original image space to a very high, possibly infinite dimensional feature space. In that high dimensional feature space, linear clustering, prediction, and classification algorithms can be applied and the results can be mapped back down to the original image space. Thus, highly nonlinear structure in the image can be recovered through the use of well-known linear mathematics in the feature space. This process has a number of advantages over traditional methods in that it allows for nonlinear interactions to be modelled with only a marginal increase in computational costs. In this paper, we present the theory of Mercer Kernels, describe its use in image mining, discuss a new method to generate Mercer Kernels directly from data, and compare the results with existing algorithms on data from the MODIS (Moderate Resolution Spectral Radiometer) instrument taken over the Arctic region. We also discuss the potential application of these methods on the Intelligent Archive, a NASA initiative for developing a tagged image data warehouse for the Earth Sciences.

Keywords: Mercer Kernels, Image Understanding, Image Analysis

1. Introduction

Currently, the Earth Observing satellites are generating multi- and hyper-spectral data at an extraordinary rate, with some systems generating data at the rate of 100 Gigabytes per hour. This data contains information vital to the understanding of the Global Ecosystem at many temporal and spatial scales. However, the vast majority of the data is archived without ever being analyzed or understood. Many scientists take extremely small samples, most of the time comprising only a few hundred or thousand images for analysis. Traditional data mining methods such as clustering and classification have been applied to image understanding problems with good success [4, 1]. In the results section below, we show the performance of a neural network to represent a traditional data mining method, and also the output of a support vector machine, which is a kernel method.

This paper addresses the problem of automatically mining the multispectral images using Mercer Kernels with the hope of finding a method to automatically generate tags for images that indicate the percentage of cloud cover, the percentage of presence of other geophysical processes such as snow, ice, melting regions, drought regions, and fire hazard.

We begin by giving a brief introduction to kernel functions. *It is important to note that the kernel methods discussed here have little relation to the established notions of kernel*

Dr. Srivastava is with the Research Institute of Advanced Computer Science, NASA Ames Research Center, and Dr. Oza is with NASA Ames Research Center.

density estimation (such as mixtures of Gaussians, Parzen windows, etc.) The hyperspectral spatiotemporal random function $Z_t(\alpha, \beta, \lambda)$ represents a series of length T of three dimensional data cubes of size $(n \times n \times \Lambda)$, where n denotes the number of pixels in one direction (assuming square images, without loss of generality), Λ denotes the total number of measured wavelengths, and T denotes the total number of time samples [2]. [1] One interpretation of a kernel function views them as a similarity metric. A common measure of similarity between two spatial locations (α, β) and (α', β') at a given time τ_0 is captured in the $(n \times n)$ linear covariance matrix $\mathbf{K}_{\tau_0}(\alpha, \beta, \alpha', \beta') = Cov(Z_{\tau_0}(\alpha, \beta, \lambda), Z_{\tau_0}(\alpha', \beta', \lambda))$. To generalize this notion of covariance, one can introduce a highly nonlinear function Φ that maps data from the Λ dimensions to a high (possibly infinite) dimensional feature space[3]: $\Phi : \mathcal{R}^\Lambda \mapsto \mathcal{H}$. Thus, we can rewrite the above covariance matrix in terms of the mapped data as follows:

$$\begin{aligned}
\mathbf{K}^{\Phi}_{\tau_0}(\alpha, \beta, \alpha', \beta') &= Cov(\Phi(Z_{\tau_0}(\alpha, \beta, \lambda)), \Phi(Z_{\tau_0}(\alpha', \beta', \lambda))) \\
&= \frac{1}{\Lambda} \sum_{\lambda=1}^{\Lambda} [\Phi(Z_{\tau_0}(\alpha, \beta, \lambda)) - \Phi(m_{\tau_0}(\lambda))][\Phi(Z_{\tau_0}(\alpha', \beta', \lambda)) - \Phi(m_{\tau_0}(\lambda))]^T \forall \alpha, \beta = 1...n.
\end{aligned}$$

Once a mapping Φ is prescribed, one can perform *linear* operations in the feature space and map the results back to the original Λ dimensional space. An interesting and extremely valuable area of research is to design the mapping Φ from the data itself and from domain knowledge. Φ does not need to be determined explicitly, since in the above calculation, only inner products arise. A kernel function is defined as the inner product of the mapped data in the feature space. We will address the problem of designing kernel functions by expressing the similarity between two samples of data using knowledge of the physics of the domain. *Unlike most other machine learning methods, kernel methods offer a unified framework to encode knowledge of the underlying physics in clustering, classification, and regression, which are common machine learning tasks.* We have already conducted research using kernel methods for discovering snow, ice, and clouds in multispectral images with very promising preliminary results (see Figures belo). In the full paper, we will discuss the Mixture Density Mercer Kernels, which allow for a Bayesian probabilistic model for the hyperspectral data using a mixture distribution with prior probabilities, thus encoding the physics of the domain into the model. The likelihood of a hyperspectral signal z_i is given by $P(z_i|\mathcal{M})$ where $\mathcal{M}$ denotes the model. A kernel function can be constructed by taking the following product of likelihoods, which encodes an independence assumption: $K(z_i, z_j) = P(z_i|\mathcal{M})P(z_j|\mathcal{M})$. **Existing Kernels** We will extensively test the existing set of kernels to determine the degree of improvement of our method compared to existing methods.

2. Image Segmentation over Snow and Ice

In this section, we describe the performance these algorithm on a real-world image segmentation problem. We obtained MODIS level 1B data for the Greenland ice sheet from the NASA Langley DAAC and mapped the data to a 1.25 km equal-area scalable Earth-grid (EASE-grid) using software developed by NSIDC to process MODIS level 1B data and convert the visible channel data to top-of-the-atmosphere (TOA) reflectances. Next the TOA reflectances were normalized by the cosine of the solar zenith angle. Only the first 7 MODIS channels were used for this study.

[1] We maintain the notion of temporal data to explicitly describe the fact that the methods discussed here can explicitly model time-dependent behavior.

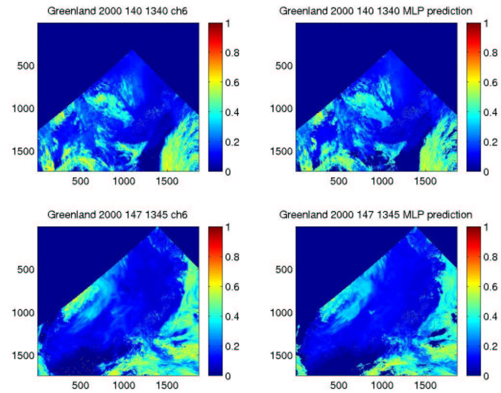

FIGURE 1. (Upper Left) This figure shows the output of Channel 6 for day 140 in the year of 2000 from the MODIS instrument over Greenland. Clouds are characterized by regions of greater density of white. (Upper Right) Shows the prediction of a multilayer perceptron with 5 inputs and 5 hidden units on the training data in the Upper Left. The bottom left panel shows a sample test image with and the bottom right shows the performance of the multilayer perceptron on the test data.

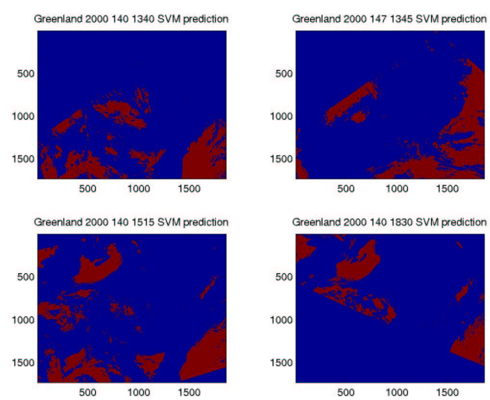

FIGURE 2. (Upper Left) This figure shows prediction of a Support Vector Machine using a Gaussian Mercer Kernel for day 140 in the year of 2000 from the MODIS instrument over Greenland. Clouds are characterized by regions of greater density of red. The remaining panels show the performance of the SVM on other test images.

3. CONCLUSIONS

A direct comparison of the traditional neural network (multilayer perceptron) model and the Mercer Kernel Support Vector Machine indicates that the kernel method may give better predictions which are more robust over areas where detection of geophysical processes is difficult, such as the detection of clouds over snow and ice.

4. Acknowledgements

The authors thank Julienne Stroeve for valuable discussions and feedback. Dr. Stroeve also provided the MODIS data used as examples in this paper. This work was supported by the NASA Intelligent Data Understanding segment of the Intelligent Systems Program.

References

1. V. Kumar, M. Steinbach, P.N. Tan, S. Klooster, and C. Potter, *Mining scientific data: Discovery of patterns in the global climate system*, Proceedings of the 2001 Joint Statistical Meeting (2002).
2. P. C. Kyriakidis and A. G. Journel, *Geostatistical space-time models: A review*, Mathematical Geology **31** (1999), no. 6, 651–684.
3. B. Scholkopf, A. Smola, and K.R. Muller, *Nonlinear component analysis as a kernel eigenvalue problem*, Neural Computation **10** (1998), 1299–1319.
4. M. Steinbach, P.N. Tan, V. Kumar, S. Klooster, and C. Potter, *Data mining for the discovery of ocean climate indicies*, Proceedings of the SIAM Conference on Data Mining, 2002 (2002).